CompCytogen 10(4): 697–717 (2016)
doi: 10.3897/CompCytogen.v10i4.11069
http://compcytogen.pensoft.net

RESEARCH ARTICLE

COMPARATIVE Cytogenetics
A peer-reviewed open-access journal
International Journal of Plant & Animal Cytogenetics, Karyosystematics, and Molecular Systematics

Chromosomal and mitochondrial diversity in *Melitaea didyma* complex (Lepidoptera, Nymphalidae): eleven deeply diverged DNA barcode groups in one non-monophyletic species?

Elena A. Pazhenkova[1,2], Vladimir A. Lukhtanov[1,2]

1 *Department of Karyosystematics, Zoological Institute of Russian Academy of Sciences, Universitetskaya nab. 1, 199034 St. Petersburg, Russias* **2** *Department of Entomology, St. Petersburg State University, Universitetskaya nab. 7/9, 199034 St. Petersburg, Russia*

Corresponding author: *Vladimir A. Lukhtanov* (lukhtanov@mail.ru)

Academic editor: *I. Gavrilov-Zimin* | Received 5 November 2016 | Accepted 21 November 2016 | Published 6 December 2016

http://zoobank.org/50FB7115-FB3A-46A7-9800-448B2B9FA3B2

Citation: Pazhenkova EA, Lukhtanov VA (2016) Chromosomal and mitochondrial diversity in *Melitaea didyma* complex (Lepidoptera, Nymphalidae): eleven deeply diverged DNA barcode groups in one non-monophyletic species? Comparative Cytogenetics 10(4): 697–717. https://doi.org/10.3897/CompCytogen.v10i4.11069

Abstract

It is generally accepted that cases of species' polyphyly in *COI* trees arising as a result of deep intraspecific divergence are negligible, and the detected cases reflect misidentifications or/and methodological errors. Here we studied the problem of species' non-monophyly through chromosomal and molecular analysis of butterfly taxa close to *Melitaea didyma* (Esper, 1779) (Lepidoptera, Nymphalidae). We found absence or low interspecific chromosome number variation and presence of intraspecific variation, therefore we conclude that in this group, chromosome numbers have relatively low value as taxonomic markers. Despite low karyotype variability, the group was found to have unexpectedly high mitochondrial haplotype diversity. These haplotypes were clustered in 23 highly diverged haplogroups. Twelve of these haplogroups are associated with nine traditionally recognized and morphologically distinct species *M. chitralensis* Moore, 1901, *M. deserticola* Oberthür, 1909, *M. didymoides* Eversmann, 1847, *M. gina* Higgins, 1941, *M. interrupta* Colenati, 1846, *M. latonigena* Eversmann, 1847, *M. mixta* Evans, 1912, *M. saxatilis* Christoph, 1873 and *M. sutschana* Staudinger, 1892. The rest of the haplogroups (11 lineages) belong to a well-known west-palaearctic species *M. didyma*. The last species is particularly unusual in the haplotypes we obtained. First, it is clearly polyphyletic with respect to *COI* gene. Second, the differentiation in *COI* gene between these mostly allopatric (but in few cases sympatric) eleven lineages is extremely high (up to 7.4%), i.e. much deeper than the "standard" DNA barcode species threshold (2.7–3%). This level of

Copyright *E.A. Pazhenkova, V.A. Lukhtanov.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

divergence normally could correspond not even to different species, but to different genera. Despite this divergence, the bearers of these haplogroups were found to be morphologically indistinguishable and, most importantly, to share absolutely the same ecological niches, i.e. demonstrating the pattern which is hardly compatible with hypothesis of multiple cryptic species. Most likely such a profound irregularity in barcodes is caused by reasons other than speciation and represents an extraordinary example of intra-species barcode variability. Given the deep level of genetic differentiation between the lineages, we assume that there was a long period (up to 5.0 My) of allopatric differentiation when the lineages were separated by geographic or/and ecological barriers and evolved in late Pliocene and Pleistocene refugia of north Africa, the Iberian and Balkan Peninsulas, the Middle East and Central Asia. We discuss the refugia-within-refugia concept as a mechanism explaining the presence of additional diverged minor haplogroups within the areas of the major haplogroups. We also provide the first record of *M. gina* in Azerbaijan and the record of *M. didyma turkestanica* as a new taxon for Russia and Europe.

Keywords

Biodiversity, butterflies, *COI*, chromosome, karyotype, mitochondrial DNA, monophyly, non-monopyletic species, Nymphalidae, phylogeography, Pleistocene refugium, taxonomy

Introduction

The *Melitaea didyma* (Esper, 1779) species complex, a group of taxa close to *M. didyma* (Bryk 1940, Higgins 1941, Kolesnichenko 1999, Kolesnichenko et al. 2011) is widely distributed in the Palaearctic region. This complex exhibits a high level of individual and seasonal variation, although distinction between described taxa and between different populations in wing pattern is often unclear (Higgins 1941, 1955, Lvovsky and Morgun 2007, Oorschot and Coutsis 2014). Simultaneously these butterflies are similar in male and female genitalia structure (Higgins 1941).

The significant reviews of this complex were published by Bryk (1940), Higgins (1941, 1955), Kolesnichenko (Kolesnichenko 1999, Kolesnichenko et al. 2011), Tuzov and Churkin (2000). More recently the whole genus *Melitaea* Fabricius, 1807 was revised by Oorschot and Coutsis (2014). However, available cytogenetic (Lukhtanov and Kuznetsova 1989), morphological (Lvovsky and Morgun 2007, Kolesnichenko et al. 2011, Oorschot and Coutsis 2014) and molecular (Wahlberg and Zimmermann 2000, Lukhtanov et al. 2009, Dincă et al. 2015) data show that the *M. didyma* species complex requires a more detailed study.

Combination of molecular and cytogenetic methods is a useful tool for detecting cryptic species (Lukhtanov et al. 2015) and can be a good addition to morphological analysis for ordering complex taxonomic structures (Lukhtanov et al. 2016). In our previous paper we applied analysis of DNA barcodes to demonstrate that *M. didyma* complex is a monophyletic group and is represented by multiple deeply diverged mitochondrial DNA haplogroups (Pazhenkova et al. 2015).

In the present study we use a combination of molecular and chromosomal markers to analyse additional material collected in Armenia, Bulgaria, Georgia, Greece, Iran, Israel, Kazakhstan, Kyrgyzstan, Russia, Slovenia, Syria and Turkey, in order to reveal

taxonomic and phylogeographic structure within the *M. didyma* species complex. In our opinion, this group includes the following species: *M. didyma* Esper, 1779, *M. chitralensis* Moore, 1901, *M. deserticola* Oberthür, 1909, *M. didymoides* Eversmann, 1847, *M. gina* Higgins, 1941, *M. interrupta* Colenati, 1846, *M. latonigena* Eversmann, 1847, *M. mixta* Evans, 1912, *M. saxatilis* Christoph, 1873 and *M. sutschana* Staudinger, 1892. This complex does not include the taxa of the *M. persea* complex (*M. persea* Kollar, 1849, *M. casta* Kollar, 1849, *M. eberti* Koçak, 1980 and *M. higginsi* Sakai, 1978) and the taxa of the *M. ala* complex (*M. ala* Staudinger, 1881, *M. bundeli* Kolesnichenko, 1999, *M. kotshubeji* Sheljuzhko, 1929, *M. acraeina* Staudinger, 1886, *M. enarea* Frühstorfer, 1917, *M. ninae* Sheljuzhko, 1935 and *M. didymina* Staudinger, 1895) which were shown to be strongly diverged with respect to genitalia structure (Higgins 1941, Kolesnichenko 1999, Oorschot and Coutsis 2014) and molecular markers (Leneveu et al. 2009).

Material and methods

We studied standard *COI* barcodes (658-bp 5' segment of mitochondrial *cytochrome oxidase subunit I*). We obtained *COI* sequences from 121 specimens collected in Armenia, Bulgaria, Georgia, Greece, Iran, Israel, Kazakhstan, Kyrgyzstan, Russia, Slovenia, Syria and Turkey. DNA was extracted from a single leg removed from each voucher specimen.

Legs from 21 specimens were processed at Department of Karyosystematics of Zoological Institute of the Russian Academy of Sciences. Primers and PCR protocol are given in our previous publications (Lukhtanov et al. 2014, Pazhenkova et al. 2015). Sequencing of double-stranded product was carried out at the Research Resource Center for Molecular and Cell Technologies of St. Petersburg State University. Legs from 100 specimens of *Melitaea* were processed at the Canadian Centre for DNA Barcoding (CCDB, Biodiversity Institute of Ontario, University of Guelph) using their standard high-throughput protocol described by deWaard et al. (2008). The set of voucher specimens of butterflies is kept in the Zoological Institute of the Russian Academy of Science (St. Petersburg).

The analysis involved 265 *COI* sequences (including outgroup) (Suppl. material 1). Among them there were 144 published sequences (Wahlberg and Zimmermann 2000, Vila and Bjorklund 2004, Leneveu et al. 2009, Lukhtanov et al. 2009, Dincă 2011, 2015, Hausmann et al. 2011, Ashfaq et al. 2013, Pazhenkova et al. 2015) collected from GenBank.

Within the studied samples, we are not completely sure of the identity of *M. chitralensis* specimens (their barcodes were obtained from GenBank) because we were not able to check these vouchers and used the identification of these samples accepted in Ashfaq et al. (2013). According to Kolesnichenko (1999), *M. chitralensis* is a member of the *M. ala* subgroup, but the analysed samples clearly clustered with *M. mixta*. Therefore, we can not exclude the possibility that these samples represent a north Pakistani population close to *M. mixta*, but not a true *M. chitralensis*.

Sequences were aligned using BioEdit software (Hall 1999). Mean uncorrected p-distances between haplogroups were calculated in MEGA7 (Kumar et al. 2015). Phylogenetic hypotheses were inferred using Bayesian inference (BI) as described previously (Vershinina and Lukhtanov 2010, Talavera et al. 2013a,b). Briefly, Bayesian analyses were performed using the program MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001) with default settings as suggested by Mesquite (Maddison and Maddison 2015): burn-in=0.25, nst=6 (GTR + I +G). Two runs of 10 000 000 generations with four chains (one cold and three heated) were performed. Chains were sampled every 10000 generations.

Karyotypes were obtained from fresh adult males and processed as previously described (Vershinina et al. 2015). Briefly, gonads were removed from abdomen and placed to freshly prepared fixative (3:1; 96% ethanol and glacial acetic acid) directly after capturing butterfly in the field. Testes were stored in the fixative for 1 month at +4°C. Then the gonads were stained in 2% acetic orcein for 7-10 days at +18-20°C. Haploid chromosome numbers (n) were counted in meiotic metaphase I (MI) and metaphase II (MII).

Results

Karyotype

The haploid chromosome number n=28 was found in prometaphase I, MI and MII cells of seven studied individuals (Table 1, Fig. 1). All chromosome elements formed a gradient size row. The karyotype contained no exceptionally large or small chromosomes.

Table 1. Chromosome number and localities of *Melitaea gina* samples collected in Iran (province West Azerbaijan) (Collectors: V. Lukhtanov, E. Pazhenkova and N. Shapoval).

Sample	Karyotype	Haplotype	Locality	Altitude	Date
Q153	n=28	M18	25 km E of Mahabad (vic. Darman): N36°45'00"; E45°51'37"	1900–2000 m	10 August 2016
Q155	n=28		25 km E of Mahabad (vic. Darman): N36°45'00"; E45°51'37"	1900–2000 m	10 August 2016
Q156	n=28	M14	25 km E of Mahabad (vic. Darman): N36°45'00"; E45°51'37"	1900–2000 m	10 August 2016
Q157	n=28	M15	25 km E of Mahabad (vic. Darman): N36°45'00,30"; E45°51'36,60"	1900–2000 m	10 August 2016
Q182	n=28		25 km E of Mahabad (vic. Darman): N36°45'00"; E45°51'37"	1900–2000 m	10 August 2016
Q183	n=28		25 km E of Mahabad (vic. Darman): N36°45'00"; E45°51'37"	1900–2000 m	10 August 2016
Q211	n=28		3 km W of Khalifen: N36°44'35"; E45°32'13"	2100–2200 m	11 August 2016

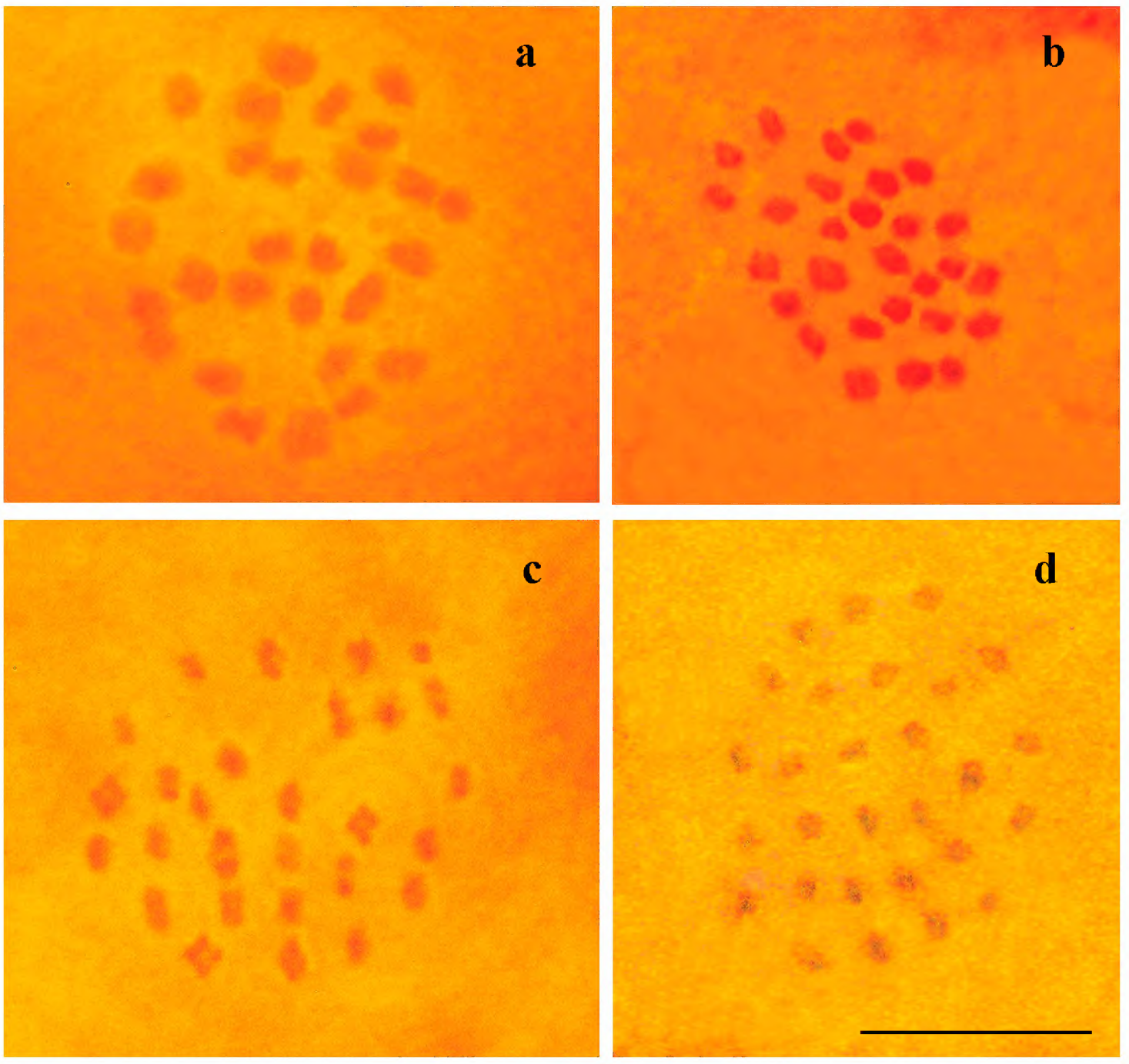

Figure 1. Karyotypes in male meiosis of *Melitaea gina* from Iran. **a** sample Q183, prometaphase I, n = 28 **b** sample Q153, late prometaphase I, n = 28 **c** sample Q183, MI, n = 28 **d** sample Q155, M I, n = 28. Scale bar corresponds to 10μ in all figures.

COI haplotypes and haplogroups

Bayesian analysis of the barcode region recovered the *M. didyma* complex as a monophyletic clade (Fig. 2), which agrees with Leneveu et al. (2009). Despite low karyotype variability, the clade was found to have unexpectedly high mitochondrial haplotype diversity. These haplotypes were clustered in 23 highly diverged haplogroups called *chitralensis*, *deserticola*, *didyma*, *didymoides*, *gina*, *gina2*, *interrupta*, *latonigena*, *liliputana*, *mauretanica*, *mixta*, *neera*, *neera2*, *occidentalis*, *protaeoccidentis*, *saxatilis*, *sutschana*, *sutschana2*, *sutschana3*, *turkestanica*, *turkestanica2*, *turkestanica3* and *turkestanica4* (Figs 2–6, Suppl. material 1). These haplogroups had high support (Bayesian posterior probability from 0.95 to 1) and were associated with particular geographical areas (Fig. 7).

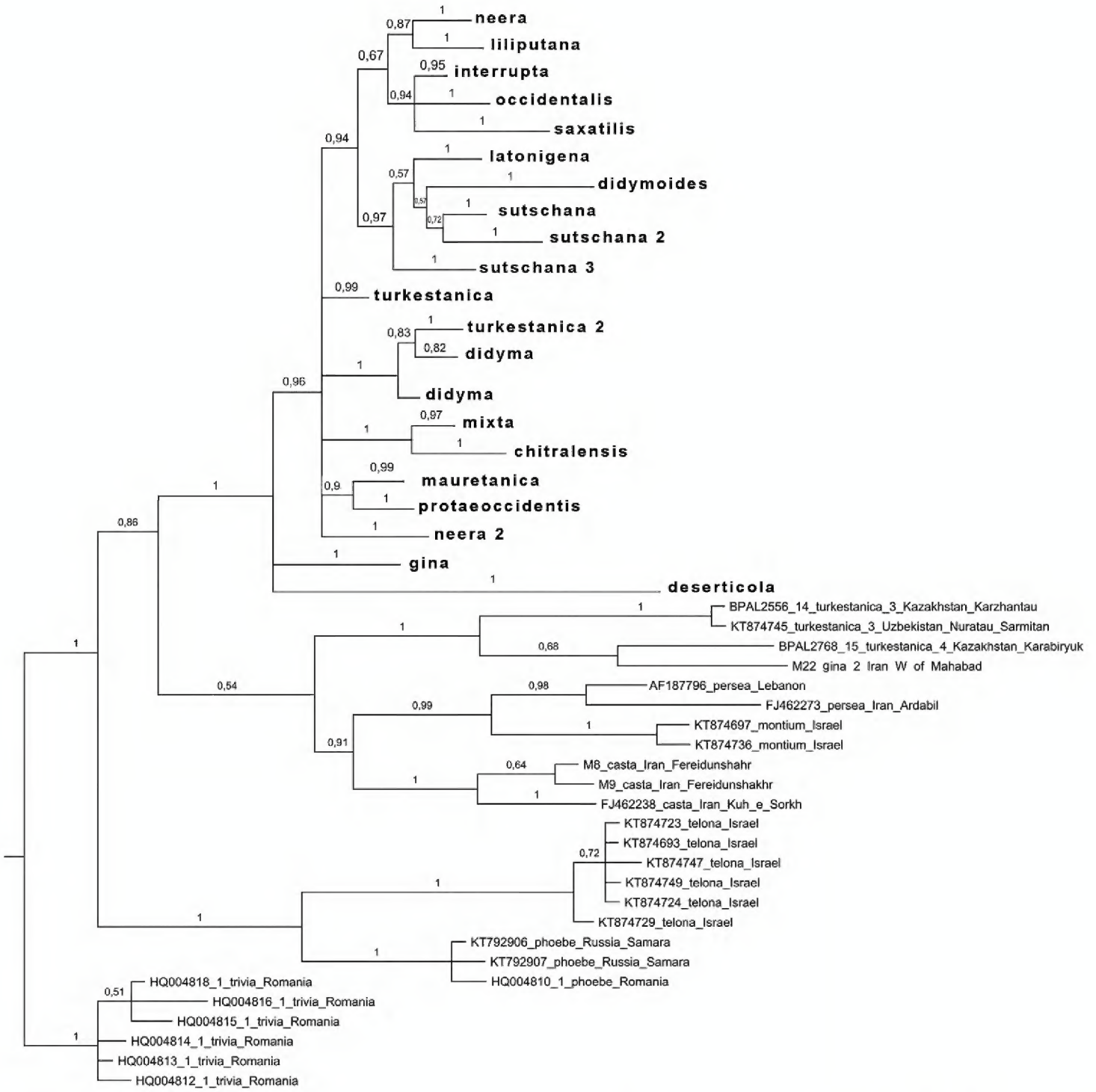

Figure 2. The Bayesian tree of *Melitaea* based on analysis of *the cytochrome oxidase subunit I* (*COI*) gene. Numbers at nodes indicate Bayesian posterior probability.

The uncorrected mean *p*-distances between the haplogroups were high (up to 9.1% between *turkestanica4* and *deserticola*) (Table 2). The majority of them were much higher than the 'standard' 2.7–3.0% DNA barcode threshold usually used for allopatric taxa as an indicator for their species distinctness (Lambert et al. 2005, Lukhtanov et al. 2015).

Most of the haplogroups were found to be allopatric. However, in some cases barcodes' clusters did not correspond to the simple allopatric geographical distribution. The sample *Melitaea gina* M22 (haplogroup *gina2*) was found in sympatry with the haplogroup *gina* in north-west Iran. The distance between *gina* and *gina2* was 6.5%. Haplogroups *turkestanica4*, *turkestanica3* and *turkestanica2* were highly diverged (up to 7.4%) as compared with the haplogroup *turkestanica* and were found in sympatry

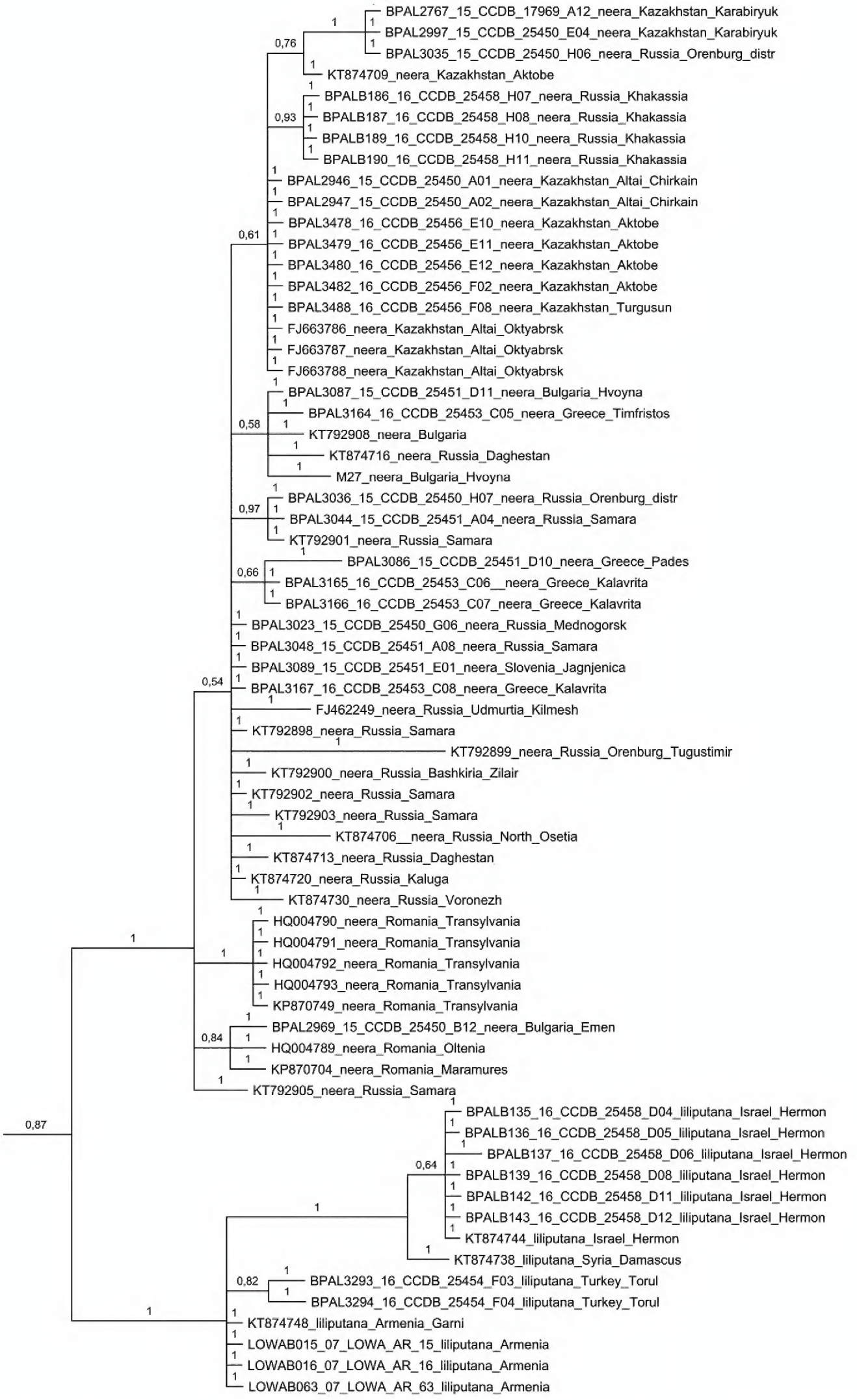

Figure 3. Fragment of the Bayesian tree of *Melitaea didyma* complex (haplogroups *neera* and *liliputana*) based on analysis of *COI* gene. Numbers at nodes indicate Bayesian posterior probability.

Table 2. Mean uncorrected *COI* p-distances between 23 haplogroups of the *M. didyma* species complex (%).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. chitralensis																						
2. deserticola	8.4																					
3. didyma	4.2	6.7																				
4. didymoides	6.3	6.9	4.7																			
5. gina	5.5	7.5	4.2	5.3																		
6. gina 2	6.4	9.5	6.5	7.2	6.5																	
7. interrupta	4.9	6.2	2.7	3.1	4.4	6.1																
8. latonigena	5.0	6.9	3.1	4.1	4.7	6.5	3.5															
9. liliputana	4.7	7.1	3.1	4.8	5.2	6.8	3.4	3.7														
10. mauretanica	4.1	6.3	2.1	4.2	4.8	7.1	2.2	3.1	3.5													
11. mixta	2.4	6.9	3.5	5.0	5.2	6.9	4.1	4.6	4.3	3.6												
12. neera	3.7	6.8	2.4	4.3	4.9	7.1	2.3	2.9	2.0	2.0	3.3											
13. neera 2	3.2	6.7	1.9	4.7	4.5	6.4	2.6	3.0	2.8	2.0	2.7	1.7										
14. occidentalis	4.9	6.9	2.9	3.9	4.6	6.4	1.8	3.8	3.9	2.4	4.0	2.8	2.4									
15. protaeoccidentis	3.3	5.6	2.1	4.1	4.2	6.7	2.7	2.7	3.0	2.1	2.9	2.1	2.0	3.0								
16. saxatilis	5.0	7.9	4.0	4.7	5.4	7.4	3.5	4.5	4.3	3.3	5.0	3.7	3.9	3.9	3.8							
17. sutschana	5.6	6.9	3.4	3.5	4.5	6.7	3.1	2.4	3.7	3.5	4.6	2.6	3.3	3.1	3.2	3.9						
18. sutschana 2	5.9	7.6	4.0	4.1	5.7	7.7	3.9	3.0	4.3	4.1	5.2	3.2	3.9	4.3	3.8	4.5	1.8					
19. sutschana 3	4.7	6.9	2.5	3.4	4.5	7.1	2.6	2.4	3.4	2.6	4.0	2.3	2.4	3.0	2.6	2.7	1.5	2.1				
20. turkestanica	3.4	7.0	2.3	4.4	4.3	7.0	3.0	3.4	3.1	2.4	2.7	2.1	1.6	3.1	2.3	3.7	3.6	4.3	2.8			
21. turkestanica 2	4.8	7.5	1.1	5.7	5.1	6.6	3.7	4.1	4.1	3.1	4.1	3.4	2.9	3.9	3.0	5.0	4.4	4.4	3.5	3.2		
22. turkestanica 3	7.0	8.9	5.8	8.9	6.4	6.4	4.8	6.1	7.1	7.0	6.3	7.9	6.9	5.8	6.0	6.2	7.0	6.9	7.0	6.6	6.4	
23. turkestanica 4	7.0	9.1	6.5	7.4	7.0	4.3	6.1	7.3	6.7	7.3	7.4	7.2	6.4	6.6	6.5	7.5	7.2	8.0	7.2	7.4	7.0	4.4

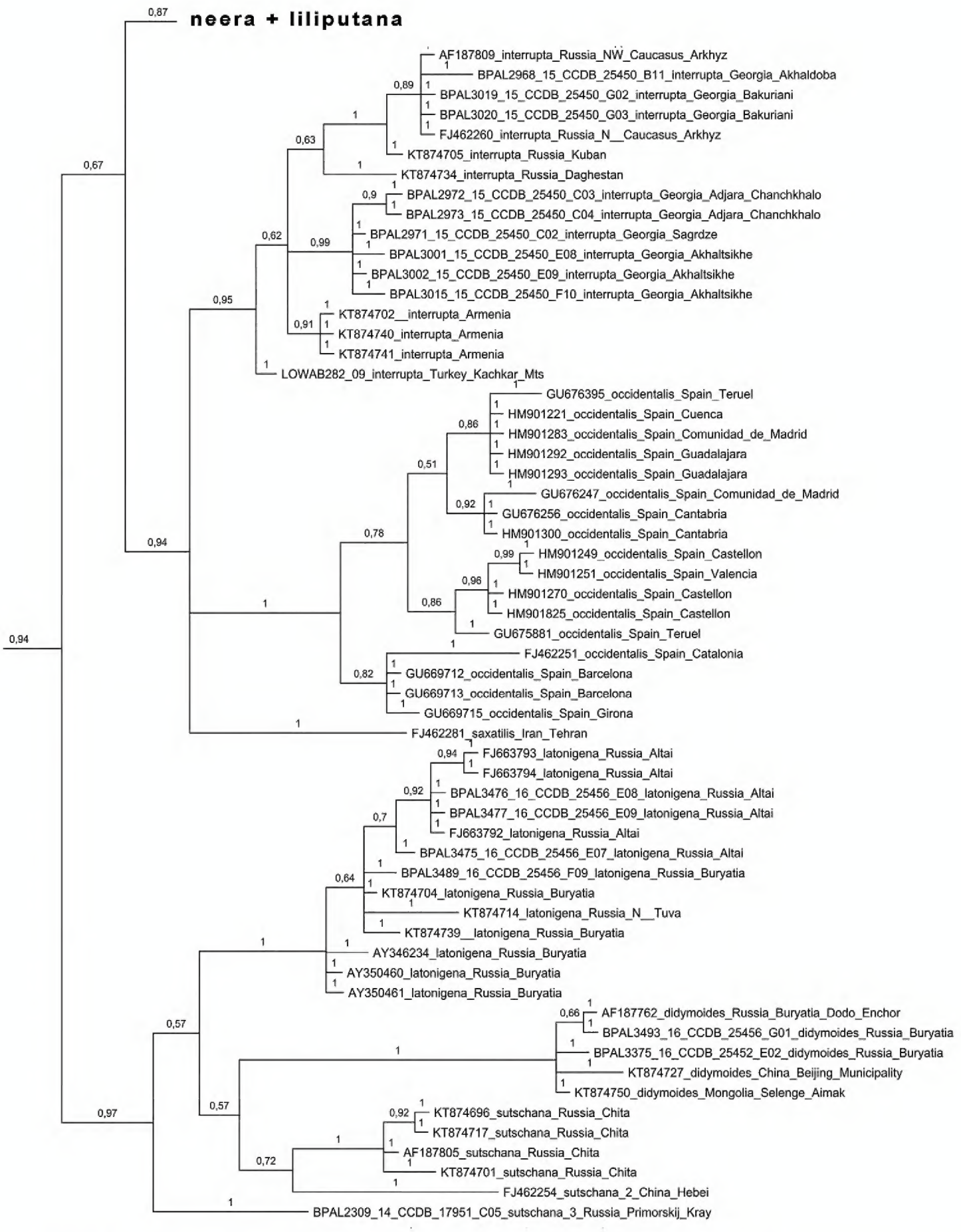

Figure 4. Fragment of the Bayesian tree of *Melitaea didyma* complex (haplogroups *interrupta*, *occidentalis*, *saxatilis*, *lathonigena*, *didymoides*, *sutschana*, *sutschana 2*, *sutschana 3*) based on analysis of *COI* gene. Numbers at nodes indicate Bayesian posterior probability.

with the haplogroup *turkestanica* (Fig. 8). In Slovenia, the specimen BPAL3090-15 (haplogroup *neera2*) was found together with the haplogroup *neera*. The distance between *neera* and *neera2* was 1.7%.

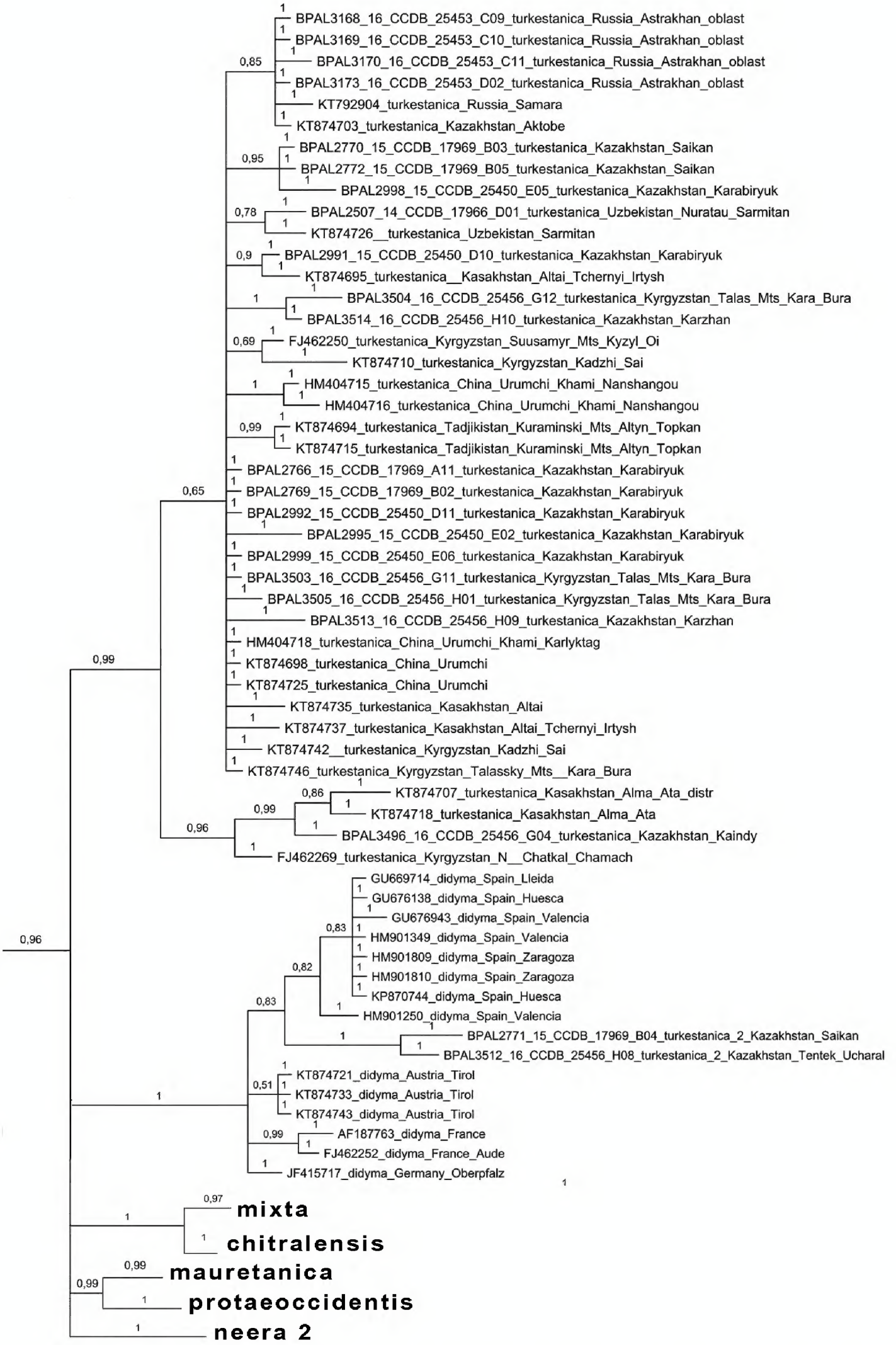

Figure 5. Fragment of the Bayesian tree of *Melitaea didyma* complex (haplogroups *turkestanica*, *turkestanica 2*, *didyma*) based on analysis of *COI* gene. Numbers at nodes indicate Bayesian posterior probability.

Two samples with the *turkestanica* haplotypes (haplogroup *turkestanica*), one from Aktobe (Kazakhstan) and one from Samara (Russia) were found in sympatry with *M. dimyma neera* haplotypes (haplogroup *neera*). In Karabiryuk (Kazakhstan), two samples with the *neera* haplotypes (haplogroup *neera*) were found in sympatry with *M. didyma turkestanica* haplotypes (haplogroup *turkestanica* and *turkestanica4*).

Discussion

Chromosome number variation

The genus *Melitaea* is known to be characterized by relatively low interspecific chromosome number variation. The representatives of basal clades (see phylogeny in Leneveu et al. 2009), the taxa of *M. cinxia* (Linnaeus, 1758), *M. diamina* (Lang, 1989), *M. athalia (*Rottemburg, 1775), *M. trivia* ([Denis & Schiffermüller], 1775) and *M. phoebe* ([Denis & Schiffermüller], 1775) species groups demonstrate n=30-31 (Federley 1938, de Lesse 1960, Robinson 1971, Larsen 1975, Hesselbarth et al. 1995). These haploid numbers are modal ones not only for *Melitaea*, but also for the family Nymphalidae and for the order Lepidoptera in whole (Robinson 1971, Lukhtanov 2000, 2014). Most likely, one of them (probably, n=31, see Lukhtanov 2014) represents an ancestral lepidopteran condition preserved in the basal lineages of *Melitaea*.

The younger lineages, the *M. fergana* Staudinger, 1882 and *M. didyma* species groups, were found to possess lower chromosome numbers varying from n=27 to n=29-30. Within the *M. fergana* species group, *M. athene* Staudinger, 1881, the only karyologically studied species, was found to have n=29 (with n=30 as a rare intra-individual variation) (Lukhtanov and Kuznetsova 1989). The species-rich *M. didyma* group consists of three complexes: a complex of taxa close to *M. ala*, a complex of taxa close to *M. persea* and a complex of taxa close to *M. didyma*. Within these complexes the following chromosome numbers were found: n=29 in *M. ala* (Lukhtanov and Kuznetsova 1989), n=27 in *M. persea* (de Lesse 1960) and different numbers from n=27 to n=29-30 in species of the *M. didyma* complex (Table 3).

Together with *M. deserticola* (n=29, Larsen 1975), *M. gina* occupies a basal position within the *M. didyma* complex (Fig. 6). Therefore analysis of *M. gina* was crucially important for understanding chromosome number evolution in this complex. Our study revealed *M. gina* to have n=28, a number previously observed in *M. didyma* from Italy (de Lesse 1960) and *M. didyma neera* from the Kazakh Altai (Lukhtanov and Kuznetsova 1989). Taking into account absence or relatively low level of interspecific chromosome number variation in the *M. didyma* complex and presence of intraspecific variation (Table 3), we conclude that in this group chromosome numbers have relatively low value as taxonomic markers (but see: Lukhtanov and Kuznetsova 1989).

Table 3. Chromosome numbers of taxa close to *M. didyma*.

Taxon	Chromosome number	Country	Locality	Reference
M. didyma ssp.	n=28	Italy	Abruzzi	de Lesse 1960
M. didyma neera	n=28	Kazakhstan	Altai	Lukhtanov and Kuznetsova 1989
M. didyma neera	n=27	Russia	N Caucasus, Pyatigorsk	Lukhtanov and Kuznetsova 1988
M. interrupta	n=29	Turkey		de Lesse 1960
M. interrupta	n=29	Azerbaijan, Nakhichevan	Zangezur Mts	Lukhtanov and Kuznetsova 1989
M. latonigena	n=29–30	Kazakhstan	Altai	Lukhtanov and Kuznetsova 1989
M. deserticola	n=29	Lebanon		Larsen 1975
M. gina	n=28	Iran	W Azerbaijan	This study

Note. We did not include in the Table 3 the following data: *M.* "*didyma*" (N Iran, Elburz, Demavend) n=28 (de Lesse 1960) because true *M. didyma* is not known from Iran (van Oorschot and Coutsis 2014), and the studied samples could represent *M. interrupta kendevana* or *M. gina*. *M.* "*didyma libanotica*" (Lebanon, Ain Zhalta Cedars) with n=27 (Larsen 1975) was also not included in the Table 2 since its identity remains unclear. The voucher samples for this count were larvae, and their identification was not certain. They actually could represent *M. persea* (n=27 is typical number for *M. persea*, including the population from Lebanon (de Lesse 1960).

DNA barcode haplogroups and problem of non-monophyletic species

Despite low level of chromosome number variability, the *M. didyma* complex was found to have unexpectedly high level of mitochondrial haplotype diversity. These haplotypes were clustered in 23 highly diverged haplogroups (Fig. 2). 12 of these haplogroups are associated with nine traditionally recognized and morphologically distinct species *M. deserticola*, *M. gina*, *M. didymoides*, *M. saxatilis*, *M. sutschana* (this species was devided recently in *M. sutschana* and *M. yagakuana* Matsumura, 1927, see Oorschot and Coutsis 2014), *M. latonigena* (this species was devided recently in *M. latonigena* and *M. latonigenides* Oorschot and Coutsis, 2014, see Oorschot and Coutsis 2014), *M. interrupta*, *M. mixta* and *M. chitralensis.*

The rest of the haplogroups belong to the well-known west-palearctic species *M. didyma*. Despite intrapopulation and seasonal variability, this species is very homogenous with respect to morphology, including the structure of genitalia, a character which is most useful for species separation in *Melitaea* (Suschkin 1913, Higgins 1941, Oorschot and Coutsis 2014). In accordance with this homogeinity, in the recent revision (Oorschot and Coutsis 2014) all populations of this species, except for Central Asian populations, were considered as members of the same subspecies *M. didyma didyma*. The populations from Central Asia were treated by Oorschot and Coutsis (2014) as a separate subspecies *M. didyma turkestanica*.

If we follow the opinion of experts in *Melitaea* taxonomy (Kolesnichenko et al. 2011, Oorschot and Coutsis 2014) and accept the traditional taxonomic treatment of the species *M. didyma*, we should acknowledge that this species is particularly unusual in the haplotypes we obtained. First, it is clearly polyphyletic with respect to *COI* gene, and

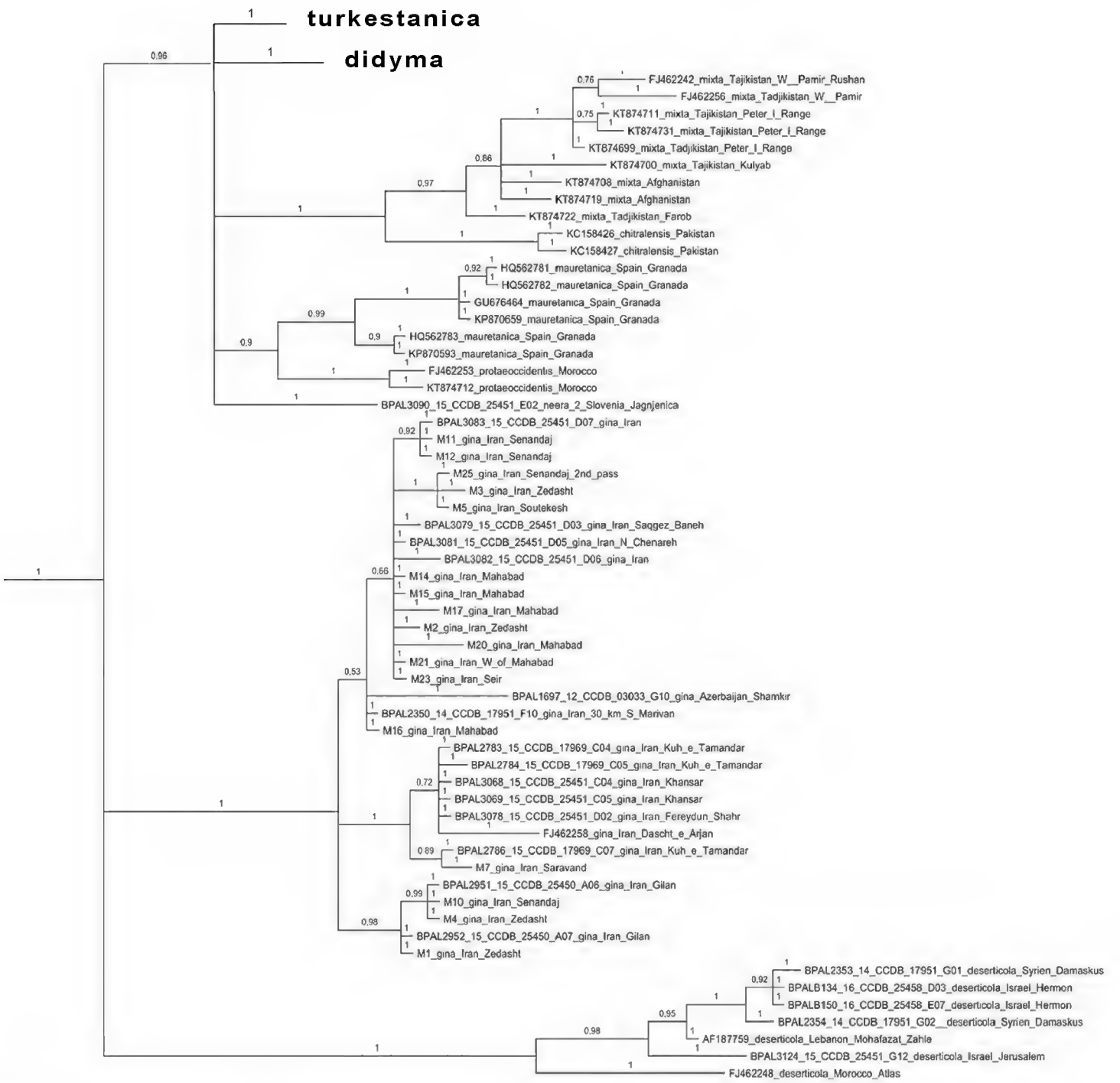

Figure 6. Fragment of the Bayesian tree of *Melitaea didyma* complex (haplogroups *mixta*, *chitralensis*, *mauretanica*, *protaeoccidentis*, *neera2*, *gina* and *deserticola*) based on analysis of the *COI* gene. Numbers at nodes indicate Bayesian posterior probability.

the lineages of *M. didyma* are intermixed with other well recognized species on the tree (Figs 2–6). Second, the number of distinct *COI* lineages within *M. didyma* is unusually high (11 lineages) and their genetic differentiation is extreme. The majority of these haplogroups are allopatric, but some of them have sympatric (*neera/neera1*, *turkestanica/turkestanica2*, *turkestanica/turkestanica3*, *turkestanica/turkestanica4*) or partially sympatric (*neera/turkestanica*, *occidentalis/didyma*) distribution. The mean uncorrected pairwise distances between the lineages is up to 7.4% if the lineages *turkestanica3* and *turkestanica4* are considered (Table 2). The lineages *turkestanica3* and *turkestanica4* are the most diverged lineages of *M. didyma*. Together with *gina2*, on the tree (Fig. 2) they have an intermediate position between the lineage (*M. didyma* + *M. deserticola* + *M. gina*) and the lineage (*M. persea* + *M. casta*). It even appears as a sister group to (*M. persea* + *M. casta*),

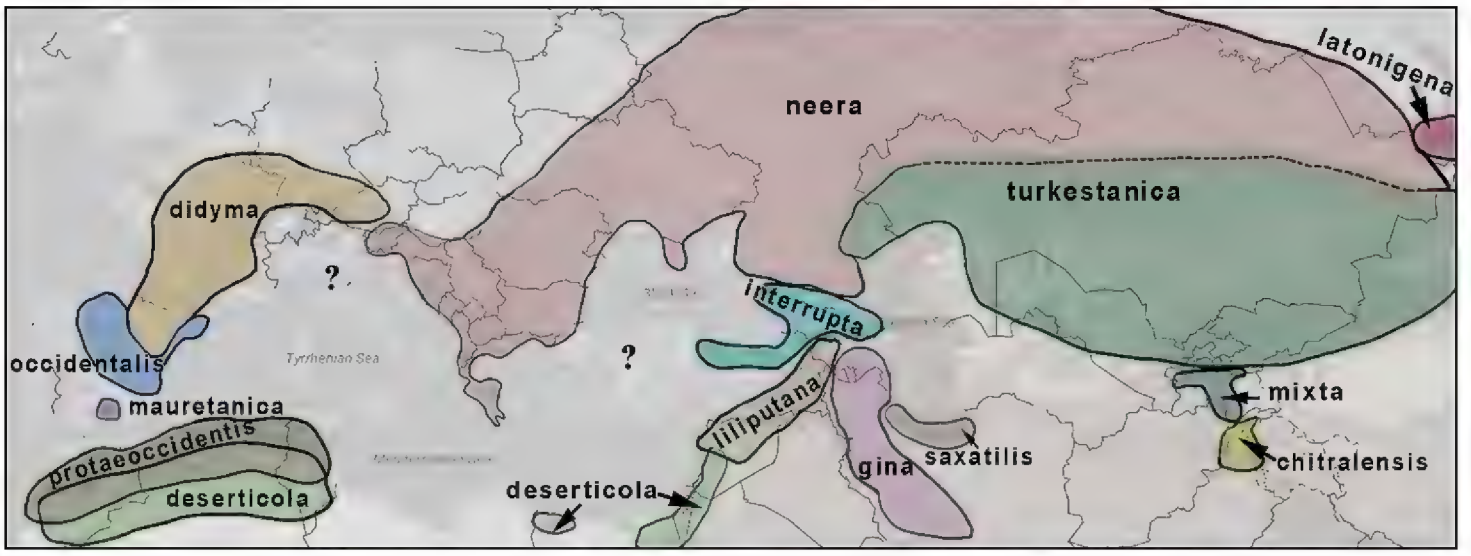

Figure 7. Distribution ranges of western *COI* haplogroups of *Melitaea didyma* complex.

Figure 8. Localization of *neera* and *turkestanica* haplogroups (yellow circles – *neera*, black – *turkestanica*, green – *turkestanica2*, red – *turkestanica3*, blue – *turkestanica4*)

but with a very low support (0.54). However, even if the lineages *gina2*, *turkestanica3* and *turkestanica4* are not considered, the distances between *M. didyma* groupings remains high, up to 4.1% between *turkestanica2* and *liliputana*, i.e. much deeper than the "standard" DNA barcode species threshold (2.7-3%) (Hebert et al. 2003, Lukhtanov et al. 2016).

There are two theoretically possible explanations for this pattern. First, *M. didyma* sensu auctorum can be a mix of multiple species that mostly have allopatric distribution ranges, but some of them are sympatric. Second, the recovered haplogroups (at least the allopatric ones) can represent highly diverged intraspecific lineages. Of course, a combination of the first and the second hypotheses is possible, and a part of the haplogroups could represent different species, and another part of the haplogroups could represent intraspecific variations.

In our opinion, the second hypothesis seems to be more plausible. There are the following arguments for the second scenario. First, no morphological differences between the bearers of these haplogroups are known (except for lighter, more yellowish wing colour in the three *M. didyma turkestanica* haplogroups as compared with other haplogroups). The second (and the most convincing) argument is based on our field obseravtion of butterfly habitats and ecological preferences. In ecology the competitive exclusion principle, also known as Gause's law is one of the most important rule (Gause 1934, Hardin 1960). In complete accordance with this rule, in case of sympatry the most closely related species pairs, such as *M. didyma*/*M. interrupta*, *M. didyma*/*M. latonigena* and *M. gina*/*M. saxatilis* demonstrate clear niche differentiation (*M. didyma* and *M. gina* are more xerophilous, whereas *M. interrupta*, *M. latonigena* and *M. saxatilis* are more mesophilous taxa). This was not a case for sympatric haplogroups *neera*/*neera2*, *turkestanica*/*turkestanica*, *turkestanica*/*turkestanica3* and *turkestanica*/*turkestanica4* (Fig. 8). The bearers of these haplogroups were not only morphologically identical, but also were found to fly exactly syntopically and synchronously. This pattern is hardly compatible with non-conspecifity of these haplogroups.

M. didyma neera and *M. didyma turkestanica* are differentiated ecologically (Pazhenkova et al. 2015), however, there was no ecological separation between bearers of the *neera* and *turkestanica* haplogroups in cases of their sympatry. In Samara and Aktobe, where the haplogroup *neera* was predominant, both haplogroups were found in *M. didyma neera* biotope (steppe), and in Karabiryuk where the haplogroup *turkestanica* was predominant, both haplogroups were found in *M. didyma turkestanica* biotope (desert) (Fig. 8). This pattern corresponds more to a result of haplotype introgression than to co-habitation of two ecologically differentiated species.

Interestingly, the haplogroup *turkestanica2* is not related to the haplogroup *turkestanica* and is a derivative from West-European haplogroup *didyma*. This pattern can be treated as a result of ancient introgression. Generally, footprints of ancient and more recent introgression are both an evidence for transparency of boundaries between *M. didyma* populations.

The mega-analysis of species-level para- and polyphyly in DNA barcode gene trees was recently conducted by using a huge data set (4977 species and 41,583 specimens of European Lepidoptera) (Mutanen et al. 2016), however without in-depth-analyses of particular cases. This study resulted in conclusion that cases of species' polyphyly in *COI* trees arising as a result of deep intraspecific divergence were negligible, and the detected cases reflected misidentifications or/and methodological errors. Despite this, our analysis demonstrates that species-level polyphyly in DNA barcode based on deep intraspecific divergence may be a real phenomenon.

Distribution ranges and phylogeography

The *M. didyma* complex consists of at least 23 *COI* haplogroups, the majority of which demonstrated a strict attachment to particular geographic ranges: *chitralensis* (north Pa-

kistan); *deserticola* (north Africa, Israel, Jordan, Lebanon, Syria); *didyma* (west Europe); *didymoides* (Asian Russia, Mongolia, North China); *gina* (W Iran, Azerbaijan); *interrupta* (Caucasus, NE Turkey); *latonigena* (Asian Russia, north-east Kazakhstan, Mongolia, north-west China); *liliputana* (Armenia, Turkey, Syria, Lebanon, Israel); *mauretanica* (south Spain); *mixta* (Tajikistan, Kyrgyzstan, Uzbekistan, Pakistan, Afghanistan); *neera* (east Europe, north Caucasus, west Siberia, north Kazakhstan); *occidentalis* (Spain); *protaeoccidentis* (north Africa); *saxatilis* (north Iran); *sutschana* (Russian Far East, Korea, north-east China) and *turkestanica* (Kazakhstan, Kyrgyzstan, Uzbekistan, Tajikistan, west China). With few exceptions (e.g. *deserticola*/*protaeoccidentis*, *deserticola*/*liliputana*), the ranges of these haplogroups do not overlap substantially (Fig. 7), and we hypothesize that mitochondrial diversity was formed in allopatry. Given the deep level of genetic differentiation between the lineages, we assume that there was a long period of allopatric differentiation when the lineages were separated by geographic or/and ecological barriers. Under generally accepted maximum 2.3% (Brower 1994) and minimum 1.5% uncorrected pairwise distance per million years (Quek et al. 2004) for *COI* sequence of various arthropod taxa, this period can be estimated to be as long as 0.5–5.0 My. In our opinion, this is an evidence that each of these haplogroups evolved in one of the main west-palaearctic late Pliocene and Pleistocene refugia in north Africa (*protaeoccidentis*, *deserticola*), the Iberian Peninsula (*occidentalis*, *mauretanica*), the Balkan Peninsula (*neera*), the Middle East (*liliputana*, *saxatilis*, *gina*) and Central Asia (*turkestanica*, *mixta*, *chitralensis*). The presence of additional diverged minor haplogroups *neera2*, *turkestanica2*, *turkestanica3*, *turkestanica4*, *gina2*, which could originate allopatrically in small isolated spots, but currently exist in secondary sympatry with major haplogroups *neera*, *turkestanica* and *gina*, agrees well with the refugia-within-refugia concept (Gòmez and Lunt 2007, Karaiskou et al. 2014). Interestingly, the area of the most diverged haplogroup *turkestanica3* is close to the area of the recently described subspecies *M. didyma carminea* (Kolesnichenko et al. 2011).

Taxonomic interpretation

We tentatively suggest interpreting the main clusters discovered within *M. didyma* sensu stricto (*M. didyma didyma*, *M. didyma mauretanica*, *M. didyma occidentalis*, *M. didyma protaeoccidentis*, *M. didyma liliputana*, *M. didyma neera* and *M. didyma turkestanica*) as subspecies because each of them has its own distribution range and is distinct with respect to mtDNA (i.e. represents by a monophyletic lineage or a combination of two or three monophyletic lineages). As a result we propose the following classification:

M. didyma **(Esper, [1779])**
- ***M. didyma didyma*** **(Esper, [1779])**
- ***M. didyma mauretanica*** **Oberthür, 1909**
- ***M. didyma occidentalis*** **Staudinger, 1961**
- ***M. didyma protaeoccidentis*** **Verity, 1929**

***M. didyma liliputana* Oberthür, 1909**
***M. didyma neera* Fischer de Waldheim, 1840**
***M. didyma turkestanica* Sheljuzhko, 1929**
***M. didymoides* Eversmann, 1847**
***M. sutschana* Staudinger, 1892**
***M. latonigena* Eversmann, 1847**
***M. interrupta* Colenati, 1846**
***M. mixta* Evans, 1912**
***M. chitralensis* Moore, 1901**
***M. deserticola* Oberthür, 1909**
***M. saxatilis* Christoph, 1873**
***M. gina* Higgins, 1941**

New records

We provide the first record of *M. gina* in Azerbaijan (sample BPAL1697-12, Azerbaijan, Shamkir, 27 June 2011, collector V. Tikhonov).

We also record *M. didyma turkestanica* as a new taxon for Russia and Europe (samples BPAL3168-16, BPAL3169-16, BPAL3170-16, BPAL3173-16 Russia, Astrakhanskaya oblast, Bogdinsko-Baskunchaksky zapovednik, 24 May 2008, collector S. Nedoshivina).

Acknowledgements

The financial support for this study was provided by the grant N 14-14-00541 from the Russian Science Foundation to the Zoological Institute of the Russian Academy of Sciences. We thank Andrei Sourakov and Andrew Warren (University of Florida) for their help in work with Lepidoptera collection in McGuire Center for Lepidoptera and Biodiversity. We are grateful to S. Nedoshivina for samples from Astrakhanskaya oblast, V.V. Tikhonov for samples from Caucasus and to N.A. Shapoval for samples from Khakassia (Russia). We thank A. Novikova for help in collecting material in Israel and N.A. Shapoval for help in collecting material in Iran. The work was partially performed using equipment of the 'Chromas' Core Facility and Centre for Molecular and Cell Technologies of St. Petersburg State University.

References

Ashfaq M, Akhtar S, Khan AM, Adamowicz SJ, Hebert PDN (2013) DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. Molecular Ecology Resources 13: 832–843. https://doi.org/10.1111/1755-0998.12131

Brower AVZ, De Salle R (1998) Patterns of mitochondrial versus nuclear DNA sequence among nymphalid butterflies: The utility of wingless as a source of characters of phylogenetic inference. Insect Molecular Biology 7: 73–82. https://doi.org/10.1046/j.1365-2583.1998.71052.x

Bryk F (1940) Geographische Variabilität von *Melitaea didyma* (Esper). Folia Zoologica et Hydrobiologica 10: 293–353.

de Lesse H (1960) Spéciation et variation chromosomique chez les Lépidoptères Rhopalocères. Annales des Sciences Naturelles (Ser. 12) 2: 1–223.

deWaard JR, Ivanova NV, Hajibabaei M, Hebert PDN (2008) Assembling DNA barcodes: analytical protocols. In: Martin CC (Ed.) Environmental Genomics, Methods in Molecular Biology. Humana Press, Totowa, New Jersey, 410: 275–283. https://doi.org/10.1007/978-1-59745-548-0_15

Dincă V, Zakharov EV, Hebert PDN, Vila R (2011) Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. Proceedings of the Royal Society B: Biological Sciences 278: 347–355. https://doi.org/10.1098/rspb.2010.1089

Dincă V, Montagud S, Talavera G, Hernández-Roldán H, Munguira ML, García-Barros E, Hebert PDN, Vila R (2015) DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. Scientific Reports 5: 12395. https://doi.org/10.1038/srep12395

Federley H (1938) Chromosomenzahlen finnländischer Lepidopteren. I. Rhopalocera. Hereditas 24: 397–464. https://doi.org/10.1111/j.1601-5223.1938.tb03219

Gause GF (1934) The struggle for existence. Baltimore, 163 pp. https://doi.org/10.5962/bhl.title.4489

Gòmez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. Phylogeography of Southern European refugia. In: Weiss S, Ferrand N (Eds) Evolutionary perspectives on the origins and conservation of European biodiversity. Springer, Amsterdam, 155–188. https://doi.org/10.1007/1-4020-4904-8_5

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

Hardin G (1960) The competitive exclusion principle. Science 131: 1292–1297. https://doi.org/10.1126/science.131.3409.1292

Hausmann A, Haszprunar G, Segerer AH, Speidel W, Behounek G, Hebert PDN (2011) Now DNA-barcoded: the butterflies and larger moths of Germany. Spixiana 34(1): 47–58.

Hesselbarth G, Oorchot H, Wagener S (1995) Die Tagfalter der Türkei unter Berücksichtigung der angrenzenden Länder. Selbstverlag Siegbert Wagener, Bocholt, Vol. 1–3, 1354 pp.

Higgins LG (1941) An illustrated catalogue of the Palearctic *Melitaea* (Lep. Rhopalocera). Transactions of the Royal Entomological Society of London 91: 175–365. https://doi.org/10.1111/j.1365-2311.1941.tb01045.x

Higgins LG (1955) A descriptive catalogue of the genus *Mellicta* Billberg (Lepidoptera: Nymphalidae) and its species, with supplementary notes on the genera *Melitaea* and *Euphydryas*. Transactions of the Royal Entomological Society of London 106: 1–131. https://doi.org/10.1111/j.1365-2311.1955.tb01262.x

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17: 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Karaiskou N, Tsakogiannis A, Gkagkavouzis K, Operator of Parnitha National Park, Papika S, Latsoudis P, Kavakiotis I, Pantis J, Abatzopoulos TJ, Triantaphyllidis C, Triantafyllidis A (2014) Greece: a Balkan subrefuge for a remnant red deer (*Cervus elaphus*) population. Journal of Heredity 105(3): 334–344. https://doi.org/10.1093/jhered/esu007

Kolesnichenko KA (1999) A review of the East Palaearctic taxa of the *Melitaea didyma* (Esper, [1779])-group. 1. The *M. ala* Staudinger, 1881 – *M. chitralensis* Moore, 1901 – complex. Atalanta 30: 87–117, 388–391.

Kolesnichenko KA, Churkin SV, Berdiev RK (2011) A review of the East Palaearctic taxa of the *Melitaea didyma* (Esper, [1779]) – group. Part II. Neue Entomologische Nachrichten 67: 257–278.

Kumar S, Stecher G, Tamura K (2015) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33(7): 1870–1874. https://doi.org/10.1093/molbev/msw054

Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PDN, Millar CD (2005) Is a large-scale DNA-based inventory of ancient life possible? Journal of Heredity 96(3): 279–284. https://doi.org/10.1093/jhered/esi035

Larsen TB (1975) Chromosome numbers and notes on testicular morphology of some Lebanese Rhopalocera (Insecta Lepidoptera). Entomologica Scandinavica 6(3/4): 253–260. https://doi.org/10.1163/187631275X00091

Leneveu J, Chichvarkhin A, Wahlberg N (2009) Varying rates of diversification in the genus *Melitaea* (Lepidoptera: Nymphalidae) during the past 20 million years. Biological Journal of the Linnean Society 97: 346–361. https://doi.org/10.1111/j.1095-8312.2009.01208.x

Lukhtanov VA (2000) Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta). Journal of Zoological Systematics and Evolutionary Research 38(2): 73–79. https://doi.org/10.1046/j.1439-0469.2000.382130.x

Lukhtanov VA (2014) Chromosome number evolution in skippers (Lepidoptera, Hesperiidae). Comparative Cytogenetics 8(4): 275–291. https://doi.org/10.3897/CompCytogen.v8i4.8789

Lukhtanov VA, Kuznetsova VG (1989) The analysis of the karyotype variability in the butterflies of the *Melitaea didyma* group along with the evidence of the species distinctness of *M. latonigena* (Lepidoptera, Nymphalidae). Zoologichesky Zhurnal 68(12): 38–46.

Lukhtanov VA, Shapoval NA, Dantchenko AV (2014) Taxonomic position of several enigmatic *Polyommatus* (*Agrodiaetus*) species (Lepidoptera, Lycaenidae) from Central and Eastern Iran: insights from molecular and chromosomal data. Comparative Cytogenetics 8(4): 313–322. https://doi.org/10.3897/CompCytogen.v8i4.8939

Lukhtanov VA, Sourakov A, Zakharov EV (2016) DNA barcodes as a tool in biodiversity research: testing pre-existing taxonomic hypotheses in Delphic Apollo butterflies (Lepidoptera, Papilionidae). Systematics and Biodiversity 14(6): 599–613. https://doi.org/10.1080/14772000.2016.1203371

Lukhtanov VA, Dantchenko AV, Vishnevskaya MS, Saifitdinova AF (2015) Detecting cryptic species in sympatry and allopatry: analysis of hidden diversity in *Polyommatus* (*Agrodiaetus*)

butterflies (Lepidoptera: Lycaenidae). Biological Journal of the Linnean Society 116(2): 468–485. https://doi.org/10.1111/bij.12596

Lukhtanov VA, Sourakov A, Zakharov EV, Hebert PDN (2009) DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. Molecular Ecology Resources 9: 1302–1310. https://doi.org/10.1111/j.1755-0998.2009.02577.x

Lvovsky AL, Morgun DV (2007) Butterflies of the Eastern Europe. KMK Scientific Press, Ltd., Moscow, Russia, 443 pp. [In Russian]

Maddison WP, Maddison DR (2015) Mesquite: a modular system for evolutionary analysis. Version 3.04. http://mesquiteproject.org

Mutanen M, Kivelä SM, Vos RA, Doorenweerd C, Ratnasingham S, Hausmann A, Huemer P, Dincă V, van Nieukerken EJ, Lopez-Vaamonde C, Vila R, Aarvik L, Decaëns T, Efetov KA, Hebert PD, Johnsen A, Karsholt O, Pentinsaari M, Rougerie R, Segerer A, Tarmann G, Zahiri R, Godfray HC (2016) Species-level para- and polyphyly in DNA barcode gene trees: strong operational bias in European Lepidoptera. Systematic Biology. https://doi.org/10.1093/sysbio/syw044

Oorschot H van, Coutsis J (2014) The Genus *Melitaea* Fabricius, 1807 (Lepidoptera: Nymphalidae, Nymphalinae). Taxonomy and systematics with special reference to the male genitalia. Tshikolovets Publications, Pardubice, Czech Republic, 360 pp.

Pazhenkova EA, Zakharov EV, Lukhtanov VA (2015) DNA barcoding reveals twelve lineages with properties of phylogenetic and biological species within *Melitaea didyma* sensu lato (Lepidoptera, Nymphalidae). ZooKeys 538: 35–46. https://doi.org/10.3897/zookeys.538.6605

Quek SP, Stuart JD, Itino T, Pierce NE (2004) Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in Crematogaster (Formicidae: Myrmicinae) inhabitants of Macaranga (Euphorbiaceae). Evolution 58: 554–570. https://doi.org/10.1111/j.0014-3820.2004.tb01678.x

Robinson R (1971) Lepidoptera genetics. Pergamon Press, Oxford, 687 pp.

Suschkin P (1913) Zur anatomischen Begründung einiger palaearktischer Arten der Gattung *Melitaea* F. (Rhopal., Nymphal.). Zeitschrift für wissenschaftliche Insektenbiologie 9: 169–175, 285–289, 321–325.

Talavera G, Lukhtanov VA, Rieppel L, Pierce NE, Vila R (2013a) In the shadow of phylogenetic uncertainty: the recent diversification of *Lysandra* butterflies through chromosomal change. Molecular Phylogenetics and Evolution 69: 469–478. https://doi.org/10.1016/j.ympev.2013.08.004

Talavera G, Lukhtanov VA, Pierce NE, Vila R (2013b) Establishing criteria for higher-level classification using molecular data: the systematics of *Polyommatus* blue butterflies (Lepidoptera, Lycaenidae). Cladistics 29: 166–192. https://doi.org/10.1111/j.1096-0031.2012.00421.x

Tuzov VK, Churkin S (2000) Genus *Melitaea*. In: Tuzov VK (Ed.) Guide to the Butterflies of Russia and Adjacent Territories (Lepidoptera, Rhopalocera). Volume 2, Pensoft Publishers, Sofia-Moscow, 59–76.

Vershinina AO, Lukhtanov VA (2010) Geographical distribution of the cryptic species *Agrodiaetus alcestis alcestis*, *A. alcestis karacetinae* and *A. demavendi* (Lepidoptera, Lycaenidae) revealed by cytogenetic analysis. Comparative Cytogenetics 4(1): 1–11. https://doi.org/10.3897/compcytogen.v4i1.21

Vershinina AO, Anokhin BA, Lukhtanov VA (2015) Ribosomal DNA clusters and telomeric (TTAGG)n repeats in blue butterflies (Lepidoptera, Lycaenidae) with low and high chromosome numbers. Comparative Cytogenetics 9(2): 161–171. https://doi.org/10.3897/CompCytogen.v9i2.4715

Vila R, Bjorklund M (2004) The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). Journal of Molecular Evolution 58(3): 280–290. https://doi.org/10.1007/s00239-003-2550-2

Wahlberg N, Zimmermann M (2000) Pattern of phylogenetic relationships among members of the tribe *Melitaeini* (Lepidoptera: Nymphalidae) inferred from mitochondrial DNA sequences. Cladistics 16: 347–363. https://doi.org/10.1006/clad.2000.0136

Supplementary material I

Table S1

Authors: Elena A. Pazhenkova, Vladimir A. Lukhtanov

Data type: Microsoft Office Excel file

Explanation note: List of *Melitaea* samples used in this study.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.